Metagenomic Next-Generation Sequencing in Clinical Microbiology

For more than a century, detection of bacterial pathogen was only achieved by obtaining pure cultures using traditional culturing methods. Increasing incidence of 'hard to treat' infections with high levels of resistance rates pose significant problems in current diagnosis and treatment options.[1] Thanks to molecular diagnostic approaches that has transformed the pathogen detection in clinical microbiology laboratories by facilitating rapid and accurate diagnosis. Clinical microbiology laboratories across the globe are in the midst of a diagnostic revolution due to emerging technologies.[2] The implementation of molecular diagnostic tools including DNA sequence-based analyses has radically altered the approach for pathogen detection. However, most molecular methods target only a selected number of pathogens using specific primers or probes.[3] The next-generation sequencing (NGS)-based approaches have been a great success in elucidating evidence to understand human pathogens at a level never possible before.[4] However, the challenges including high cost, complexity and expertise required in sequence analysis have hindered the implementation of NGS-based techniques in resource-limited clinical laboratories.[5]

In the past few years, the introduction of low-cost sequencing platforms and the availability of commercial software solutions suggest the integration of NGS in clinical laboratories.[6] With increasing applications of metagenomics in the present-day clinical microbiology laboratories, it is important for clinical microbiologists to fully understand the diagnostic implications of clinical metagenomics. This editorial provides an opportunity for clinical microbiologists to understand the promises and implementation problems of metagenomic NGS (mNGS) in diagnostic laboratories. Comparative analysis of mNGS with currently practised sequencing methods expected to provide a better understanding about the clinical utility of obtained data. Multiple case reports from both bacteriology and virology laboratories can help in learning how mNGS can be implemented in real-world clinical settings.

Overview of Metagenomic Next-Generation Sequencing

The clinical mNGS approach emerged as a promising alternative to traditional diagnostic techniques since it can directly characterise nucleic acids of all potential pathogen, including viable but uncultivable microorganisms present in a sample.[7] Sequencing of a metagenome generates thousand to billions of reads in a single run of all potential pathogens, including bacteria, viruses, fungi and parasites. Metagenomics approaches can carry out pathogen identification, outbreak investigations and resistance gene mapping in clinical microbiology laboratories with high accuracy.[8] Furthermore, sequencing directly from clinical specimens can provide complete diagnostic information in time-efficient manner. The development of metagenomic analysis in clinical laboratories has changed the microbiologists from asking not only 'Who is there?' but also other important questions such as 'What are they doing?' and 'How did they reach here?'.[9]

mNGS methods are generally classified into two approaches based on the target or the kind of information obtained while sequencing. The polymerase chain reaction (PCR) amplification based targeted metagenomic sequencing, includes 16S rRNA gene or ITS-based approaches that provide only taxonomical classification for broad range of pathogens. On the other hand, shotgun-based approaches provide unlimited accesses to other clinically relevant genomic features such as resistance gene profiling, virulence factors detection and strain-level typing of all microbial communities in the sample.[10-12] Hence, these two approaches need to be selected according to the information required for the intended purpose. In theory, both metagenomic approaches found to be ideal for clinical laboratories to handle with the increasing complexity of pathogens. Yet, many challenges still remain for the practical integration of metagenomics into the clinical microbiology workflow.[13] Currently, a number of mNGS workflows are available for the implementation of clinical metagenomics. The optimisation of mNGS protocols requires the user to identify critical steps as well as the challenges posed by currently established workflows.

Metagenomics Workflow

The proposed mNGS workflow of a clinical microbiology laboratory for the identification of clinically important pathogens is shown in Figure 1. Metagenomics study design involves four major steps as follows: (i) sample/specimen collection, handling and storage, (ii) nucleic acid extraction, (iii) library preparation and sequencing and (iv) bioinformatics analysis.[10] The most critical steps in mNGS are the extraction and sequencing protocols. The choice of technical factors involved in the selected extraction and sequencing protocols can impact the overall output of the experiment.[12]

Despite the initial interests, mNGS still has many bottlenecks to overcome. Major bottlenecks that need to be resolved include (i) lower concentration of microbial genomic content in a clinical specimen, (ii) overwhelming amount of human host DNA present in the clinical specimen and (iii) lack of established protocols for the laboratory validation of

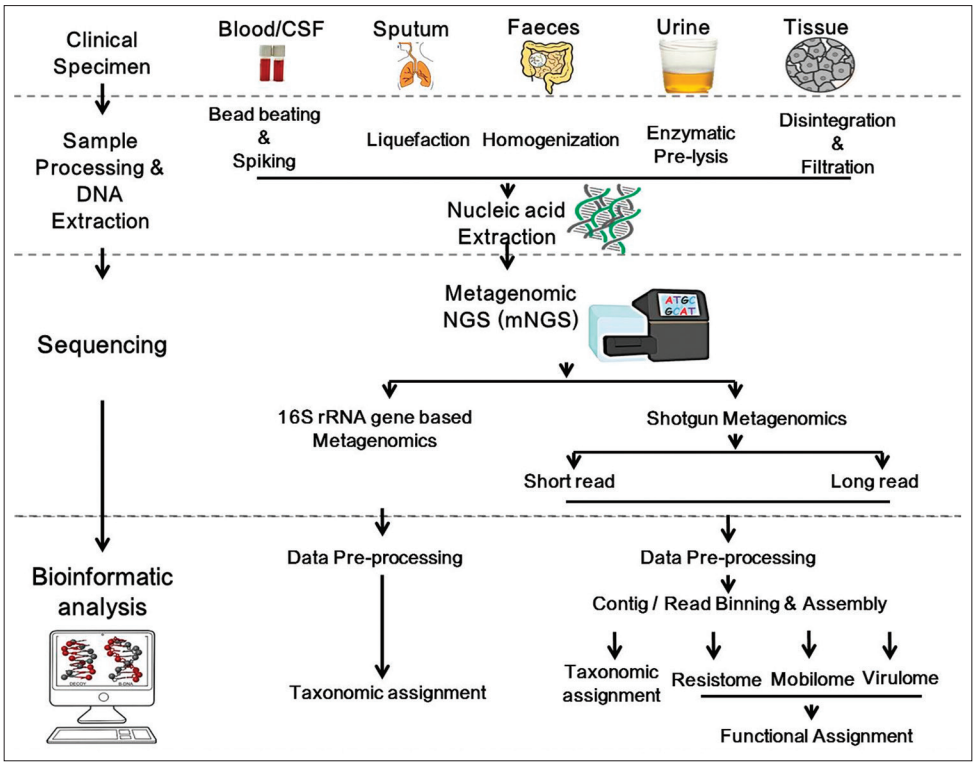

Figure 1: Schematic representation depicting the standard metagenomic workflow

mNGS assays.[14,15] In addition, the difficulty associated with understanding the latest advances from concepts to possible applications by clinicians or microbiologists found to be a considerable barrier. A better understanding of the on-going advances in nucleic acid extraction, sequencing technology and computational biology by clinical microbiologists can be the first step in overcoming these technical limitations.

Sampling and Nucleic Acid Extraction

The advantage of applying mNGS in diagnostic laboratories can be completely exploited by steps such as unbiased sampling, standard sample storage/transport and extracting high-quality nucleic acid.[9] Biases in these steps can influence the outcome of a metagenomic study. The processing of clinical sample for bacterial and viral metagenomics should be consistent with standard microbiological practices, including aseptic handling, to prevent contamination. The storage and transport mostly vary and highly specific for certain clinical specimen types.[16] For example, disparities in the sample matrix and the concentration of target pathogens in diverse clinical samples including blood, faeces, tissues, respiratory tract secretions and urine potentially impact the quality and quantity of nucleic acid extracted.[17] In the case of faeces samples, the temperature during transportation should be at 4°C to prevent bacterial growth and thereby alteration in the microbial composition. As a result, several methodologies and workflows have been introduced to explore mNGS in recent years. In general, both faeces and sputum samples need to be liquefied, while tissue samples should be homogenised or disintegrated. Samples such as blood, cerebrospinal fluid (CSF), urine and swabs will have low microbial biomass and hence effective strategies need to be implemented for sample processing.[18]

Genomic DNA or RNA protocols need to be standardised and validated for the selected clinical specimen prior to the sequencing. An ideal nucleic acid extraction protocol should extract high concentration of target pathogens DNA without co-extracting any residual host background DNA and inhibitors.[9] As a result of high variability in the quality and quantity, nucleic acid extraction step was given particular attention generating large number of extraction protocols and commercial kits. Most of the studies in recent times have used commercial nucleic acid extraction kits, but these kits are designed for PCR ready nucleic acids.[10] Prior knowledge of the sample specimen and expected pathogen may be required for ensuring sequencing ready DNA/RNA. Further downstream experiments for host DNA depletion and/or target bacterial enrichment are advised to be combined with kit method. In particular saponin-based methods have been most effective in reducing background human DNA while improving the sensitivity of mNGS to detect pathogens. The preferred strategies for sample processing and nucleic acid extraction are shown in Table 1.

Sequencing Platforms

For more than three decades, Sangers sequencing (1st generation) was considered as the gold standard in clinical laboratories for DNA sequence-based diagnostic tests. Although Sanger sequencing has undergone significant modifications by means

of automation and analysis, the technical challenges such as limited capacity (single gene/400–800 nucleotides in single run), relatively higher operating costs and laborious protocols prevented the widespread use in clinical laboratories.[21] On the other hand, NGS has allowed simultaneous analysis of multiple genes or genomes with high-throughput and low sequencing costs. Even NGS has diversified into a number of different sequencing platforms as the technologies continue to improve.[22] Since the choice of sequencing platform depends on the intended applications, clinical microbiologist requires a basic understanding about the different commercially available NGS platforms.

The NGS era was first established by the introduction of Roche's 454 pyrosequencing technology in 2005 (2nd generation). Pyrosequencing utilised a sequence-by-synthesis (SBS) approach to sequence the fragmented DNA library after amplification by emulsion PCR.[23] Following this, the benchtop short-read sequencing platforms Illumina 'MiSeq' and Life technologies 'Ion torrent PGM' instruments are the two commonly used NGS platforms in clinical laboratories. The Illumina platforms were successful in utilising the SBS technology based on the fluorescent detection, while complementary fluorescently tagged nucleotides are incorporated. Alternatively, the Ion Torrent platform is a semiconductor-based DNA sequencing technology that detects sequence data by sensing a change in pH during template-directed DNA synthesis.[24] Despite the popularity of these platforms, their sequencing technologies are incapable to provide the complete information of a genome as they are limited to sequence small fragments and later assembled as multiple contigs.

The introduction of third-generation sequencing platforms such as Pacific BioSciences (PacBio) and Oxford Nanopore Technology (ONT) revolutionised the genome sequencing technology by establishing long linear reads (1–100 kb) in very short time (2–10 h). Both technologies do not need any pre-amplification steps and utilise single-molecule sequencing technology (SMRT). Precisely, PacBio uses the incorporation of fluorescent labelled nucleotide in SBS approach, while ONT identifies the nucleotides based on the changes in electrical charges as DNA traverses the biological nanopores available in the flow cell. Long read sequencing platforms provide full structure of microbial genome with ample coverage to distinguish between chromosomes, plasmids and other mobile genetic elements.[25]

On the whole, Sanger's sequencing technology can be preferred to obtain information on single genes. Short read platforms of Illumina and ion torrent provide highly accurate, but incomplete sequencing reads. Long read technology of PacBio and ONT achieves complete circular bacterial genomes or plasmids but less accurate. The preference of the technology in clinical microbiology laboratories need to be based on the diagnostic goal. Some of the commonly used sequencing platforms in clinical laboratories and its features are given in Table 2.

Computational Tools for Metagenomics

The bioinformatics workflows also subjected to vary according to the type of data and research goal. In general,

Table 1: Preferred strategies for sample processing and nucleic acid extraction in metagenomic next generation sequencing study

Sample type	Sample processing	Challenges	DNA extraction
Blood/CSF/sterile fluids	Bead beating and spiking	Contamination and low bacterial load	QIAamp blood DNA mini kit
Faecal	Homogenization	Complex sample with inhibitors	QIAamp stool DNA mini kit
Sputum	Liquification	Complex sample with inhibitors	Zymo quick-DNA Universal Kit[19]
Tissue	Disintegration and filtration	Host DNA	DNeasy blood and tissue kit Molzym Micro-Dx™ kit
Urine	Enzymatic prelysis	Low bacterial load and inhibitors	QIAamp DNA mini kit
Swabs	Mechanical disruption	Low bacterial load	Solid phase reversible immobilization[20]

CSF: Cerebrospinal fluid

Table 2: Commonly used sequencing platforms in clinical laboratories and its features

Platform/instrument	Methodology/chemistry	Average read length (bp)	Read type	Challenges
ABI sanger 3500 genetic analyser	PCR and chain termination	500-900	Single end	Substitution
Roche 454 GS junior	Emulsion PCR and pyrosequencing	200-400	Single end	Indels
Illumina				
MiSeq	Bridge amplification and reversible dye terminator sequencing	300	Paired end	Substitution
HiSeq		150		
Ion torrent PGM	Emulsion PCR and Ion semiconductor sequencing	200-400	Single end	Homopolymer errors
PacBio RSII	SMRT	Upto 10 kb	Single end	Indels
Oxford Nanopore MinIon	Single molecule sequencing with a membrane protein nanopore	Upto 20kb	Single end	Indels

SMRT: Single molecule real time, PCR: Polymerase chain reaction

the computational assembly tools have been developed for both targeted metagenomic sequencing (16S, 18S, ITS) and whole-genome (shotgun) metagenomic sequencing data. In addition, more than 80 tools are currently available for metagenomic sequencing analysis, and determining the best tool among the bunch is a challenging task.[26] To address the challenge, basic information related to the commonly used tools are discussed below.

The bioinformatics analysis of 16S metagenomics mainly includes preprocessing and operational taxonomic units (OTUs) assignment. There are many open-source tools available to perform these tasks. However, QIIME[27] and MOTHUR[28] are the two widely used pipelines that can provide accurate and comparable results. Preprocessing of metagenomics reads is the critical step in the 16S-based metagenomics analysis. The contaminants present in sequence reads need to be removed and low-quality bases will be filtered by quality trimming. Improper preprocessing steps affect the downstream analysis largely. Hence the total microbial population in a sample is obtained by the binning of reads and clustering on the basis of similarity referred to as OTUs. These individual OTUs are specific to each bacterial species.[29]

Shotgun metagenomics approach helps in deciphering both microbial population and also its function. A typical bioinformatics pipeline for analysing the shotgun metagenomics includes pre-processing, binning, assembly and functional annotation.[30] Quality control steps can be used to filter the low quality reads thereby preventing the misassemblies of the bacterial genome. Binning classify the metagenomics reads into its corresponding taxonomy. It can be performed either with sequencing reads or post assembly. However, binning prior to assembly can efficiently minimise the assembly process.[12] The binned reads are assembled individually to obtain single contiguous sequence. Further, the assembled genomes are annotated and predicted proteins are classified into protein families.

The introduction of third-generation long-read sequencing technologies (Pacific Biosciences and ONT) resulted in longer sequence reads (1–100 kb) in comparison to short-read technologies (Illumina and Ion torrent) that can only produce reads <500 bp. Although these cost-effective long read technologies generated much excitement in the beginning, higher error rates (5%–15%), complex workflows and lack of state of art metagenomic tools pose many challenges.[31,32] Continuous updation of sequencing chemistry and development of hybrid assembly tools combining both short and long metagenomic reads alleviated some of these challenges. Details on some of the common and emerging bioinformatic pipelines for metagenomic analysis are described in Table 3. These advanced computer algorithms have achieved high quality near-complete genomes and generating information up to subspecies level.[56]

Detection of Viral Pathogens

Conventionally, viral pathogens are detected by visualising the cytopathic effect in cell culture monolayers or by antibody neutralisation tests. However, many viral types are un-cultivable in the laboratory conditions, and antibody neutralisation tests depend on the availability of quality antiserum. Molecular methods such as PCR/quantitative real-time-PCR needs sequence information for the target viruses beforehand. The metagenomic identification of viral pathogens has gained popularity among researchers because of the unbiased detection of complete taxonomic composition in a clinical sample. Through metagenomics, uncultivable or novel viral pathogens that may be responsible for many uncommon disease etiologies can be identified.[57]

Although the recent advances in sequencing technology have improved the application of metagenomics in clinical virology laboratories, standardised/validated protocols are

Table 3: Bioinformatic tools and pipelines for clinical metagenomic analysis

Methods	Computational tools	Function	References
16S metagenomics	FastQc	Quality control checks	[33]
	FastX	Preprocessing	[34]
	AmpliconNoise	Removal of noise	[35]
	UCHIME	Detecting chimeric sequences	[36]
	Mothur and QIIME	Pipeline for 16S metagenomics analysis	[27,28]
Shotgun short read metagenomics	Kaiju and MetaPhyler	Taxonomic classification	[37,38]
	MEGAN	Analysis of large metagenomic datasets	[39]
	MetaWRAP	Metagenomics analysis pipeline	[40]
	metaSPAdes, MetaVelvet and Genovo	Assembler	[41-43]
	GeneMark, Prodigal and Glimmer-MG	Gene prediction	[44-46]
	Pfam, Go database and KEGG	Functional annotation	[47-49]
Shotgun long read metagenomics	CRuMPIT and SqueezeMeta	Automated metagenomics pipeline	[50,51]
	metaFlye	Long-read metagenome assembler	[52]
	WIMP	Real-time species identification on the MinION	[53]
	Centrifuge	Microbial classification	[54]
	CheckM	Assessing the quality of metagenomes	[55]
Hybrid metagenomics	OPERA-MS	Hybrid metagenomic assembler for long error-prone reads	[56]

currently lacking.[58] The general metagenomics workflow for the identification of viral pathogens includes the three main steps as explained previously. However, for the simultaneous detection of both DNA and RNA viruses, reverse transcription of mRNA to cDNA step need to be included before the sequencing step. Among the given steps, bioinformatics analysis is considered as the critical step in identifying viral pathogens.[59] To avoid further confusion, a set of workflow suitable for different studies were tested and described in the following review articles.[57,60]

The first few reports on the application of viral metagenomics were from environmental samples.[61,62] Interestingly, the first study on the metagenomic assessment of human virulome was conducted using faecal samples.[63] These preliminary viral metagenomic studies conducted on pre-NGS era used shearing of the community DNA followed by cloning and targeted sequencing to obtain the complete community data. After establishing protocols for DNA viruses, the RNA viruses were also detected from the human gut microflora after the addition of the reverse transcription step.[64] The clinical application of viral metagenomic drew one step close to reality when total DNA virus community from blood samples were identified by viral metagenomics in 2005.[65] Similar protocols were used to analyse the DNA virus community of other clinical specimens such as oral cavity, sputum and tissue samples.[66]

The first clinical application of viral metagenomics was in 2008 when a novel arenavirus was detected from a patient suffering from transplant associated diseases.[67] Later, a number of previously unknown and potentially pathogenic viruses were also identified using 454/Roche-based viral mNGS study with similar workflow.[66] In addition, many case reports suggest the detection of viral pathogens by mNGS when other conventional methods such as PCR have failed. Some of the major contributions of metagenomics in clinical virology include the diagnosis of fatal human cases of infectious encephalitis,[68] Zika fever,[69] Ebola[70] etc. mNGS also have identified fevers of unknown origin caused by hepatitis C virus, chikungunya virus,[71] dengue virus,[72] West Nile virus,[73] and human herpesviruses. The clinical application of viral mNGS provided key information about which therapeutic measures to develop.

Detection of Bacterial Pathogens

The complete microbial profile within a clinical sample may not be easily detectable by conventional culture methods. For instance, traditional culture-based techniques are designed to target mostly aerobic pathogens. In addition, not all bacteria can be effectively cultured from a sample containing multiple organisms. The identification and characterisation of bacterial isolates from clinical samples also depends on various factors such as growth requirements, viability of existing organisms or growth inhibition of pathogenic bacteria due to bacteriocin production.[74] On the other hand, metagenomic sequencing can detect the entire microbial genome regardless of the culture requirements and phenotypical characteristics. mNGS for bacterial detection is particularly attractive as most clinical samples constitutes mixture of different organisms with varying genome size (3–8 mb) and composition (30%–60% GC). In addition, the shotgun metagenomic sequencing of bacterial samples can provide the data on important functional capabilities such as antibiotic-resistant genes, virulence factors and mobile genetic elements. Hence, mNGS approaches, by sequencing the whole genome of all the microorganisms present in a sample, can make accurate diagnosis and treatment of infections.[7]

Similar to the viral metagenomics, standardised protocols are currently lacking particularly for the data processing and analysis for bacterial metagenomic analysis. An outline of the standard workflow for the implementation of a metagenomics approach to identify and characterise bacterial pathogens are summarised in Figure 1. As described previously, the standard mNGS pipeline for bacterial metagenomics involves four major steps: (i) sample collection, processing and transport, (ii) the extraction of inhibitor-free metagenomic DNA, (iii) library preparation and sequencing and (iv) sequence read assembly and analysis using bioinformatic analysis.[9] In a clinical setting, untargeted mNGS is perhaps the most unbiased approach for the comprehensive diagnosis and novel treatment for bacterial infections. Rapid determination of disease etiology, antimicrobial susceptibility pattern and potential pathogenicity can often assist the clinicians to make a decision and initiate appropriate therapy. Metagenomic sequencing of clinical samples after regular intervals during the therapy can have important implications for on-going treatment.[75,76] Rapid identification of pathogen, especially in meningitis, sepsis and pneumonia can improve patient outcomes and minimise the use of broad-spectrum, empiric antibiotics in therapy. A combined effort from clinicians regarding the disease etiology along with metagenomic information could help in successful implementation of mNGS in clinical settings.

The clinical application of mNGS includes the detection of pathogens in some rare or complex cases where conventional methods have failed.[77] Nearly, all mNGS workflow for bacterial infections were developed for meningitis, sepsis and pneumonia and some protocols are now available for clinical reference testing of patients.[78,79] Moreover, mNGS has also been performed in different type of clinical sample, including CSF, blood, respiratory secretions, urine, stool or tissue. Unlike viral metagenomics, reports on the successful application of mNGS in detecting bacterial pathogens are only a handful. Many proof of concept has been demonstrated in the detection of bacterial pathogens such as *Streptococcus pneumoniae*,[80] *Klebsiella pneumoniae* and *Haemophilus influenzae*[81] in patients with pneumonia. Similarly, *Mycobacterium tuberculosis* was detected in a patient with respiratory failure[82] and *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *K. pneumoniae* from patients infected with transfusion-related sepsis.[83]

Metagenomics and Antimicrobial Resistance

From the literature, it is evident that the majority of the metagenomics studies have been conducted to identify the prevalence of uncharacterised or emergent or novel microbial pathogens.[84] Although diagnostic shotgun metagenomics offers an added advantage of identifying genes that have a functional role in antibiotic resistance or virulence or other important metabolic process, most studies were microbiome studies and are fundamentally designed to characterise the microbial communities. Clinical metagenomics also includes the screening of genes that has functional role in metabolism, virulence or antimicrobial resistance from clinical samples.[85] According to the current understanding, the identification of antimicrobial resistance genes (ARGs) in clinical metagenomics can be implemented as sequence-based and functional metagenomics.[86,87]

Sequence-based identification of ARGs involves metagenomic sequencing of a selected microbial community using NGS platforms and characterisation of resistome by means of bioinformatic tools.[87] The identification and classification of ARGs from metagenomic reads are generally carried out by mapping against databases such as ResFinder,[88] comprehensive antibiotic resistance database (CARD),[89] ARG-ANNOT[90] or Resfams.[91] However, sequence-based identification of ARGs may be biased as the identification is dependent on databases. Hence, novel or uncharacterised ARGs cannot be detected. Function-based metagenomics have been emerged to challenge the threats of the new millennium such as antimicrobial resistance, virulence and mobile genetic elements. In functional metagenomics, the metagenomic library (<10 kb) fragments are cloned into vectors such as plasmids, fosmids, cosmids or bacterial artificial chromosomes.[87] The vector is then transformed into an expression host, and clones hence formed, will be screened in selective media for the presence of antibiotic-resistant clones. The ARG insert in the selected clones can be further sequenced and identified. Therefore, the detection of unknown ARGs is possible with this approach.

Several studies have investigated the human resistome by means of sequence-based metagenomic approach. A major study conducted on intestinal resistome by Forslund *et al*. analysed 252 faecal metagenomes from the USA, Denmark and Spain.[92] A substantial number of studies have been conducted on investigated the human resistome in different clinical specimens.[93-95] On the other hand, a handful of functional metagenomics studies on resistome analysis are only available in the public domain.[96-98]

Limitations and future directions

One of the major limitations of mNGS is that the sensitivity of the approach is critically dependent on the level of background. Increased human host background relative to microbial specimen resulted in a reduced number and proportion of microbial reads and hence a decrease in mNGS sensitivity. In addition, defining specific microbial profiles that are diagnostic or predictive of disease development can be difficult. The cost of mNGS based diagnostic tests as well as limited specimen for further discrepancy testing are some of the other bottlenecks in implementing the techniques in clinical laboratories. Moreover, metagenomics is a novel and rapidly developing discipline. Therefore, standardised protocols are currently lacking, especially for the data processing and analysis, which require high computational resources and bioinformatics expertise.

Although metagenomics approaches have been successful in tracking infections and outbreaks, more should be done to forecast the transmission routes and prevent the disease spread. One of the major goals for the near future is tweaking the technology towards clinical utility and validation according to clinical laboratories. For this objective, mNGS assay workflows should be standardised for the detection of multiple pathogens from various clinical specimens. Improving the quantitation and sensitivity of the current protocols to suit more towards the microbial specimen by removing interfering substances (host DNA) can be another priority. Last of all, technological advancements in handling, interpreting and making use of the tremendous amount of sequencing data are expected. This would in turn make the technology more cost-effective.

Jobin John Jacob, Balaji Veeraraghavan, Karthick Vasudevan

Department of Clinical Microbiology, Christian Medical College, Vellore - 632 004, Tamil Nadu, India.
E-mail: jjj1986@gmail.com

References

1. Li B, Webster TJ. Bacteria antibiotic resistance: New challenges and opportunities for implant-associated orthopedic infections. J Orthop Res 2018;36:22-32.
2. Messacar K, Parker SK, Todd JK, Dominguez SR. Implementation of rapid molecular infectious disease diagnostics: The role of diagnostic and antimicrobial stewardship. J Clin Microbiol 2017;55:715-23.
3. Adzitey F, Huda N, Ali GR. Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks 3 Biotech 2013;3:97-107.
4. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, *et al.* Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol 2017;243:16-24.
5. Di Resta C, Galbiati S, Carrera P, Ferrari M. Next-generation sequencing approach for the diagnosis of human diseases: Open challenges and new opportunities. EJIFCC 2018;29:4-14.
6. Kohlmann A, Grossmann V, Haferlach T. Integration of next-generation sequencing into clinical practice: Are we there yet? Semin Oncol 2012;39:26-36.
7. Greninger AL, Naccache SN. Metagenomics to assist in the diagnosis of bloodstream infection. J Appl Lab Med 2019;3:643-53.
8. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. Metagenomics for pathogen detection in public health. Genome Med 2013;5:81.
9. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. Annu Rev Pathol 2019;14:319-38.
10. Forbes JD, Knox NC, Peterson CL, Reimer AR. Highlighting clinical metagenomics for enhanced diagnostic decision-making: A step towards wider implementation. Comput Struct Biotechnol J 2018;16:108-20.
11. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet 2019;20:341-55.
12. Couto N, Schuele L, Raangs EC, Machado MP, Mendes CI, Jesus TF, *et al.* Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens. Sci Rep 2018;8:13767.
13. Balloux F, Brønstad Brynildsrud O, van Dorp L, Shaw LP, Chen H,

Harris KA, *et al.* From theory to practice: Translating whole-genome sequencing (WGS) into the clinic. Trends Microbiol 2018;26:1035-48.
14. Dekker JP. Metagenomics for clinical infectious disease diagnostics steps closer to reality. J Clin Microbiol 2018;56. pii: e00850-18.
15. Dulanto Chiang A, Dekker JP. From the pipeline to the bedside: Advances and challenges in clinical metagenomics. J Infect Dis 2019. pii: jiz151.
16. Martin TC, Visconti A, Spector TD, Falchi M. Conducting metagenomic studies in microbiology and clinical research. Appl Microbiol Biotechnol 2018;102:8629-46.
17. Zhang D, Lou X, Yan H, Pan J, Mao H, Tang H, *et al.* Metagenomic analysis of viral nucleic acid extraction methods in respiratory clinical samples. BMC Genomics 2018;19:773.
18. Wylezich C, Papa A, Beer M, Höper D. A versatile sample processing workflow for metagenomic pathogen detection. Sci Rep 2018;8:13108.
19. Terranova L, Oriano M, Teri A, Ruggiero L, Tafuro C, Marchisio P, *et al.* How to process sputum samples and extract bacterial DNA for microbiota analysis. Int J Mol Sci 2018;19. pii: E346.
20. Vo AT, Jedlicka JA. Protocols for metagenomic DNA extraction and illumina amplicon library preparation for faecal and swab samples. Mol Ecol Resour 2014;14:1183-97.
21. Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. Hugo J 2011;5:1-2.
22. Levy SE, Myers RM. Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet 2016;17:95-115.
23. Siqueira JF Jr., Fouad AF, Rôças IN. Pyrosequencing as a tool for better understanding of human microbiomes. J Oral Microbiol 2012;4:10743.
24. Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochim Biophys Acta 2014;1842:1932-41.
25. Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmler S, *et al.* Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. BMC Microbiol 2019;19:143.
26. McIntyre AB, Ounit R, Afshinnekoo E, Prill RJ, Hénaff E, Alexander N, *et al.* Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. Genome Biol 2017;18:182.
27. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335-6.
28. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75:7537-41.
29. Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol 2011;77:3219-26.
30. Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 2014;5:209.
31. Korlach J. Understanding accuracy in SMRT® sequencing. Pacific Biosci 2013;1-9.
32. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, *et al.* Assessing the performance of the oxford nanopore technologies MinION. Biomol Detect Quantif 2015;3:1-8.
33. Patel RK, Jain M. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. PLoS One 2012;7:e30619.
34. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics 2011;27:863-4.
35. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. BMC Bioinformatics 2011;12:38.
36. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;27:2194-200.
37. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 2016;7:11257.
38. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. BMC Genomics 2011;12 Suppl 2:S4.
39. Huson DH, Weber N. Microbial community analysis using MEGAN. Methods Enzymol 2013;531:465-85.
40. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 2018;6:158.
41. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. Genome Res 2017;27:824-34.
42. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: An extension of velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res 2012;40:e155.
43. Laserson J, Jojic V, Koller D. Genovo: De novo assembly for metagenomes. J Comput Biol 2011;18:429-43.
44. Besemer J, Borodovsky M. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 2005;33:W451-4.
45. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
46. Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL. Gene prediction with glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Res 2012;40:e9.
47. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, *et al.* Pfam: The protein families database. Nucleic Acids Res 2014;42:D222-30.
48. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, *et al.* The gene ontology (GO) database and informatics resource. Nucleic Acids Res 2004;32:D258-61.
49. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
50. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, *et al.* Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. BMC Genomics 2018;19:714.
51. Tamames J, Puente-Sánchez F. SqueezeMeta, A highly portable, fully automatic metagenomic analysis pipeline. Front Microbiol 2018;9:3349.
52. Kolmogorov M, Rayko M, Yuan J, Polevikov E, Pevzner P. MetaFlye: Scalable long-read metagenome assembly using repeat graphs. bioRxiv 2019.
53. Urda D, Subirats JL, García-Laencina PJ, Franco L, Sancho-Gómez JL, Jerez JM. WIMP: Web server tool for missing data imputation. Comput Methods Programs Biomed 2012;108:1247-54.
54. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res 2016;26:1721-9.
55. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;25:1043-55.
56. Bertrand D, Shaw J, Kalathiyappan M, Ng AH, Kumar MS, Li C, *et al.* Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol 2019;37:937-44.
57. Nooij S, Schmitz D, Vennema H, Kroneman A, Koopmans MP. Overview of virus metagenomic classification methods and their biological applications. Front Microbiol 2018;9:749.
58. Bhukya PL, Nawadkar R. Potential applications and challenges of metagenomics in human viral infections. In: Metagenomics for Gut Microbes. IntechOpen, London, United Kingdom; 2018. p. 19.
59. Brinkmann A, Andrusch A, Belka A, Wylezich C, Höper D, Pohlmann A, *et al.* Proficiency testing of virus diagnostics based on bioinformatics analysis of simulated *In silico* high-throughput sequencing data sets. J Clin Microbiol 2019;57. pii: e00466-19.
60. Rose R, Constantinides B, Tapinos A, Robertson DL, Prosperi M. Challenges in the analysis of viral metagenomes. Virus Evol 2016;2:vew022.
61. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, *et al.* Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U S A 2002;99:14250-5.
62. Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, *et al.* Diversity and population structure of a near-shore marine-sediment viral community. Proc R Soc Lond B Biol Sci 2004;271:565-74.
63. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, *et al.* Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 2003;185:6220-3.
64. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, *et al.* RNA

viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol 2006;4:e3.
65. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. Biotechniques 2005;39:729-36.
66. Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research. Virology 2012;434:162-74.
67. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, *et al.* A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 2008;358:991-8.
68. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: Application of next generation sequencing for undiagnosed cases. J Infect 2018;76:225-40.
69. Sardi SI, Somasekar S, Naccache SN, Bandeira AC, Tauro LB, Campos GS, *et al.* Coinfections of zika and chikungunya viruses in Bahia, Brazil, identified by metagenomic next-generation sequencing. J Clin Microbiol 2016;54:2348-53.
70. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, *et al.* Emergence of zaire ebola virus disease in guinea. N Engl J Med 2014;371:1418-25.
71. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, *et al.* Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med 2015;7:99.
72. Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Servetas SL, Mateczun AJ, *et al.* Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics 2014;15:96.
73. Wilson MR, Zimmermann LL, Crawford ED, Sample HA, Soni PR, Baker AN, *et al.* Acute West Nile virus meningoencephalitis diagnosed via metagenomic deep sequencing of cerebrospinal fluid in a renal transplant patient. Am J Transplant 2017;17:803-8.
74. Abayasekara LM, Perera J, Chandrasekharan V, Gnanam VS, Udunuwara NA, Liyanage DS, *et al.* Detection of bacterial pathogens from clinical specimens using conventional microbial culture and 16S metagenomics: A comparative study. BMC Infect Dis 2017;17:631.
75. Lim YW, Evangelista JS 3rd, Schmieder R, Bailey B, Haynes M, Furlan M, *et al.* Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. J Clin Microbiol 2014;52:425-37.
76. Wu SC, Rau CS, Liu HT, Kuo PJ, Chien PC, Hsieh TM, *et al.* Metagenome analysis as a tool to study bacterial infection associated with acute surgical abdomen. J Clin Med 2018;7. pii: E346.
77. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, *et al.* Chronic meningitis investigated via metagenomic next-generation sequencing. JAMA Neurol 2018;75:947-55.
78. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, *et al.* Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res 2019;29:831-42.
79. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol 2019;4:663-74.
80. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, *et al.* Variability and diversity of nasopharyngeal microbiota in children: A metagenomic analysis. PLoS One 2011;6:e17035.
81. Yan Q, Cui S, Chen C, Li S, Sha S, Wan X, *et al.* Metagenomic analysis of sputum microbiome as a tool toward culture-independent pathogen detection of patients with ventilator-associated pneumonia. Am J Respir Crit Care Med 2016;194:636-9.
82. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ 2014;2:e585.
83. Crawford E, Kamm J, Miller S, Li LM, Caldera S, Lyden A, *et al.* Investigating transfusion-related sepsis using culture-independent metagenomic sequencing. Clin Infect Dis 2019. pii: ciz960.
84. Oniciuc EA, Likotrafiti E, Alvarez-Molina A, Prieto M, Santos JA, Alvarez-Ordóñez A. The present and future of whole genome sequencing (WGS) and whole metagenome sequencing (WMS) for surveillance of antimicrobial resistant microorganisms and antimicrobial resistance genes across the food chain. Genes (Basel) 2018;9. pii: E268.
85. Forbes JD, Knox NC, Ronholm J, Pagotto F, Reimer A. Metagenomics: The next culture-independent game changer. Front Microbiol 2017;8:1069.
86. De R. Metagenomics: Aid to combat antimicrobial resistance in diarrhea. Gut Pathog 2019;11:47.
87. Willmann M, Peter S. Translational metagenomics and the human resistome: Confronting the menace of the new millennium. J Mol Med (Berl) 2017;95:41-51.
88. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, *et al.* Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 2012;67:2640-4.
89. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 2017;45:D566-73.
90. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212-20.
91. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 2015;9:207-16.
92. Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A, *et al.* Country-specific antibiotic use practices impact the human gut resistome. Genome Res 2013;23:1163-9.
93. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat Commun 2013;4:2151.
94. Quirós P, Colomer-Lluch M, Martínez-Castillo A, Miró E, Argente M, Jofre J, *et al.* Antibiotic resistance genes in the bacteriophage DNA fraction of human fecal samples. Antimicrob Agents Chemother 2014;58:606-9.
95. Buelow E, Gonzalez TB, Versluis D, Oostdijk EA, Ogilvie LA, van Mourik MS, *et al.* Effects of selective digestive decontamination (SDD) on the gut resistome. J Antimicrob Chemother 2014;69:2215-23.
96. Sommer MO, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. Science 2009;325:1128-31.
97. Fouhy F, Ogilvie LA, Jones BV, Ross RP, Ryan AC, Dempsey EM, *et al.* Identification of aminoglycoside and β-lactam resistance genes from within an infant gut functional metagenomic library. PLoS One 2014;9:e108016.
98. Moore AM, Patel S, Forsberg KJ, Wang B, Bentley G, Razia Y, *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. PLoS One 2013;8:e78822.

Received : 21-10-2019 **Revised :** 31-10-2019
Accepted : 05-11-2019 **Published Online :** 19-11-2019

Access this article online

Quick Response Code:

Website: www.ijmm.org

DOI: 10.4103/ijmm.IJMM_19_401

How to cite this article: Jacob JJ, Veeraraghavan B, Vasudevan K. Metagenomic next-generation sequencing in clinical microbiology. Indian J Med Microbiol 2019;37:133-40.